2005

Analysis of Job-Training Effects on Korean Women

Myoung-jae Lee
Singapore Management University, mjlee@smu.edu.sg

S. J. Lee

Citation

Lee, Myoung-jae and Lee, S. J.. Analysis of Job-Training Effects on Korean Women. (2005). *Journal of Applied Econometrics*. 20, (4), 549. Research Collection School Of Economics.
Available at: https://ink.library.smu.edu.sg/soe_research/378

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **20**: 549–562 (2005)
Published online 7 April 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.771

ANALYSIS OF JOB-TRAINING EFFECTS ON KOREAN WOMEN

MYOUNG-JAE LEE[a]* AND SANG-JUN LEE[b]
[a] *School of Economics and Social Sciences, Singapore Management University, Singapore*
[b] *Korea Research Institute for Vocational Education and Training, Seoul, South Korea*

SUMMARY

We analyse job-training effects on Korean women for the period January 1999 to March 2000, using a large data set of size about 52,000. We employ a number of estimation techniques: Weibull MLE and accelerated failure time approach, which are both parametric; Cox partial likelihood estimator, which is semiparametric; and two pair-matching estimators, which are in essence nonparametric. All of these methods gave the common conclusion that job training for Korean women increased their unemployment duration. The trainings were not cost-effective in the sense that they took too much time 'locking in' the trainees during the training span, compared with the time they took to place the trainees afterwards. Despite this negative finding, some subgroups had positive effects: white-collar workers trained for finance/insurance or information/communication.

1. INTRODUCTION

Job training has a long history for many developed countries. In Korea, job training was almost negligible before the financial crisis during 1997–1998 which drove many people out onto the street unemployed. This led to the first ever large-scale job training in Korea, and since then job training has been the mainstay in the Korean government's strategy to cope with unemployment. Given the amount of money poured into job training in many countries, the importance of job training needs no justification, and there are various methods available these days to evaluate job-training effects [see, for example, Angrist and Krueger (1999), Heckman *et al.* (1999) and references cited therein].

The goal of this paper is to assess job-training effects on Korean women by comparing a treatment group ($d = 1$) of job-trained women with a control group ($d = 0$) of women who received unemployment insurance benefit instead of job training. Specifically, we will examine whether the training shortens or not the unemployment duration. In doing so, we address three issues. First, the duration is censored for some women. Second, the training may shorten the duration for some women while lengthening it for some others, in which case we have to define what 'training effect' means. Third, with these two issues present, there are many different ways to estimate job-training effects; one would like to know which is better, or whether the different methods yield the same conclusion or not.

Let y_{ji} be the 'potential response' when the treatment $d = j$, $j = 0, 1$, is 'exogenously given to' individual i. For individual i, only one of y_{1i} and y_{0i} is observed, and the realized response variable is $y_i \equiv d_i y_{1i} + (1 - d_i)y_{0i}$. In our data, y_{ji} is unemployment duration, which is censored for those who remain still unemployed when the data survey ended. We posit that there exists a

* Correspondence to: Professor Myoung-jae Lee, School of Economics and Social Sciences, Singapore Management University, 469 Bukit Timah Road, Singapore 259756. E-mail: mjlee@smu.edu.sg

Received 14 August 2002
Revised 2 October 2003

censoring time, say c_i, and the observed response t_i is

$$t_i = \min(y_i, c_i) = \min(d_i y_{1i} + (1 - d_i)y_{0i}, c_i)$$

Although the calendar censoring timing (the data the data survey ended) is the same for all individuals, since the individuals entered the data at different time points, c_i varies across i.

Define a no-censoring indicator $q_i \equiv 1[y_i \leq c_i]$ where $1[A] = 1$ if A holds and 0 otherwise. We have data on

$$d_i, q_i, t_i, x_i, \quad i = 1, \ldots, N, \quad iid \text{ across } i \tag{1}$$

where x_i is a covariate vector. Other than x_i affecting y_{ji}, we can think of unobserved variables affecting y_{ji}, denoted as u_{ji}, $j = 0, 1$ or just u_i if $u_{1i} = u_{0i}$. Since duration happens over time, some covariates can be time-variant, but in our data, none are. In view of the *iid* assumption, we will often omit the subscript i indexing individuals. Assume that

$$c_i \text{ is independent of } d_i, x_i, y_{0i}, y_{1i} \tag{2}$$

as usually done in duration analysis. This may not hold if the macroeconomic condition changes much during the survey period, because c_i depends on the timing of job loss.

The so-called 'mean treatment effect' is

$$E(y_1 - y_0) = E(y_1) - E(y_0)$$

If d is uncorrelated with y_0 and y_1, then the mean treatment effect is the same as the 'group mean difference' $E(y|d = 1) - E(y|d = 0)$. The condition $\mathrm{COR}(d, y_j) = 0$, $j = 0, 1$, holds easily if d is randomized as in clinical trials. In our data, d is not randomized but self-selected by individuals. Thus, the treatment group ('T group') can be different at a pre-treatment stage from the control group ('C group'), and the difference may be in x or in u_0 and u_1. This difference should be eliminated; otherwise, we may mistake the effects of these differences as the effects of the treatment difference. The difference in x may cause an 'overt bias', and the difference in u_0 or u_1 may cause a 'hidden bias'. Most effort in treatment effect analysis with observational data is spent on eliminating these biases.

Overt bias can be removed by conditioning on x: instead of the unconditional group mean difference, look at the conditional group mean difference

$$E(y|x, d = 1) - E(y|x, d = 0) = E(y_1|x, d = 1) - E(y_0|x, d = 0) = E(y_1|x) - E(y_0|x)$$

where the last equality holds if

$$y_j,\ j = 0, 1, \text{ are independent of } d \text{ given } x \tag{3}$$

This is a 'no-selection-bias' assumption, which assumes away hidden bias. Without this assumption, essentially, we need an instrument for d to deal with hidden bias. But instruments do not come easy, and we do not have any plausible instrument for d in our data. The no-selection-bias assumption will be maintained throughout this paper.

Once the conditional effect $E(y_1 - y_0|x)$ is identified, we can get an x-weighted average ('marginal effect') $\int E(y_1 - y_0|x)\omega(x)dx$, where $\omega(x)$ is a weighting function. The natural choice for $\omega(x)dx$ is $dF(x)$, where F is the distribution function for x.

When can we declare job training a success? There could be many different opinions on this question, but we can think of at least two criteria: unemployment duration and post-training wage. Unfortunately, in our data, post-training wage comes from the survey respondents' recollection, which is unreliable, while pre-training wage comes from the tax authority, which is reliable. Thus, in this paper, we will examine only unemployment duration. Focusing on unemployment duration only, a large proportion (71%) of which is censored in our data, we can think of a couple of different ways to characterize the training success, and the estimation method differs depending on which is looked at. We show three different measures of training success that will be used in this paper.

First, as a variation of the mean effect, consider the 'proportional effect'

$$E\{\ln(y_1) - \ln(y_0)|x\} = E\{\ln(1 + (y_1 - y_0)/y_0)|x\} \simeq E\{(y_1 - y_0)/y_0|x\} \tag{4}$$

Using $\ln(y_1)$ and $\ln(y_0)$ instead of y_1 and y_0 gives a proportional measure of treatment effect, getting rid of the scale in the response variable. The scale is also removed if $E(y_1 - y_0|x)/E(y_0|x)$ is used, as in Lee and Kobayashi (2001). But $E\{\ln(y_1) - \ln(y_0)|x\}$ has an additional attraction: we can analyse duration in an 'accelerated failure time framework', where ln(duration) is specified as a linear function of d and x; if one follows the more popular 'hazard-based' duration approaches, estimating equation (4) would not be straightforward.

Under censoring, the accelerated failure time approach requires specifying the distribution function for ln(duration) given x; we will use the normal distribution. For the sake of comparison, two well-known hazard-based approaches, the Weibull maximum likelihood estimator and the Cox partial likelihood estimator, will be used as well. The latter is semiparametric in that it does not fully specify the distribution of duration given x.

Second, instead of running the risk of misspecifying the model to estimate equation (4) under censoring, we may nonparametrically estimate

$$\begin{aligned} &E\{\ln(y)|x, d = 1, q = 1\} - E\{\ln(y)|x, d = 0, q = 1\} \\ &= E\{\ln(y_1)|x, y_1 \le c\} - E\{\ln(y_0)|x, y_0 \le c\} \end{aligned} \tag{5}$$

under conditions (2) and (3); the condition $q = 1$ is to select only the non-censored observations. To control for x nonparametrically, we will do 'matching', which will be reviewed in section 4. In equation (5), censored observations are not used, which is a loss of information.

When we say 'no effect', it can mean many different things; e.g., when median or quantile effects are used as in Lee (2000), no effect means no median or quantile effects. In equation (4), no effect means no change in E(ln(response)), which can happen even if $y_0 \neq y_1$. The most strict version of no effect requires $y_0 = y_1$. Under this strict version, equation (5) becomes zero. But when $y_0 \neq y_1$ is allowed under no effect, equation (5) may not be zero, and is hard to interpret differently from equation (4) due to the conditioning events $y_1 \le c$ and $y_0 \le c$.

Third, as just noted, equation (5) has two shortcomings of ignoring the censored observations and being hard to interpret. To overcome these problems, instead we may take

$$P(y_1 \le \tau|x) > P(y_0 \le \tau|x) \text{ for some } \tau \tag{6}$$

as a measure of training success for the sub-population characterized by x. Imagine two groups of people with the same x who entered the data at the same time, and one group is treated and the other not. Let τ be the span between the data-entering time point and the survey-ending time point.

Matching on x and the data-entering time point t_e, the inequality (6) can be nonparametrically estimated where τ is the study-ending time point minus t_e. Since t_e and x vary across individuals, (6) can be computed for different τ and x, and a weighted average

$$\int \{P(y \leq \tau|x, d = 1) - P(y \leq \tau|x, d = 0)\} w(\tau, x) d\tau dx$$

may be used as a marginal effect where $w(\tau, x)$ is a weighting function.

In clinical trials, subjects drop out because they are cured or because they perceive the treatment ineffective; also possible is dropout caused by the treatment 'pain/burden' relative to the subject's patience. In our data, there are also dropouts from the job-training programmes. The reasons for the dropouts, however, are not known in the data. Due to this problem, dropouts will not be used in our analysis.

Section 2 describes the details of our data. Section 3 reviews the duration analysis methods for equation (4), and provides our empirical findings. Section 4 reviews matching and applies it to equations (5) and (6). Finally, Section 5 concludes.

2. DATA DESCRIPTION

Our data consist of two sub-files from the Center for Employment Information in the Department of Labor in South Korea: the job-training file for the T group and the unemployment-insurance file for the C group. The C group is not a 'no-training' group; rather, they are the unemployed who chose to receive unemployment insurance. Both files include only women with some records on unemployment insurance premium payment; i.e., workers at firms/factories without ever paying for unemployment insurance premium are not covered in the files. Among the women who became unemployed in 1999, those who took job training and completed it by the end of 1999 and those who received unemployed insurance benefit which ended by the end of 1999 are in our data set; there are $N_1 = 5031$ treated units and $N_0 = 47{,}060$ control units. They are then followed up until the end of March 2000; this short follow-up, decided by the Department of Labor, resulted in high censoring percentages (71% for the T group and 72% for the C group).

Although an unemployed woman can choose d to some extent, she has to meet some criteria to be eligible for unemployment insurance: paying for the insurance at least for 0.5 years is required. For those ineligible women, self-selection for d is less worrisome because $d = 0$ is ruled out by law. Unemployment insurance does not last for more than six months in general, although there are exceptions depending on factors such as age and disability. Also, if one quits her job voluntarily, she is not eligible for unemployment insurance in principle. The trainees receive some compensation, the amount of which varies across the trainees. Roughly, the training compensation is about \$300 per month, and the unemployment insurance benefit is the minimum of about \$700 and 50% of the last workplace monthly wage. Thus, strictly speaking, the treatment of interest in our data is a combination of job training, no unemployment insurance and training compensation. That the C group receive substantially more money is likely to induce the C group to stay unemployed longer than necessary [e.g., Meyer (1990), Røed and Zhang (2003)]; if anything, our finding will be biased favourably for the T group.

Suppose we compare two women, one treated and one control. Matching on the quit/fired and data-entering dates as well as on other covariates does two things: first, the two women searched for jobs equally long before joining the study (and thus were equally frustrated); second, since they entered the data at the same time, we can compare them with 1[employed by 31/3/2000]. With this response variable, the censoring poses no problem, because we know whether the women are employed or not by 31/3/2000 regardless of the censoring. The training duration is included in the unemployment duration; that is, a trainee is regarded as unemployed.

There are many different types of job training, of varying durations in days, as shown in Table I. There were in fact more types of job training, but those with only a few trainees were removed from the data. The definition of 'service industry' is not clear in the data source.

The aforementioned institutional facts along with self-selection led to visible differences in the two groups as shown in the summary statistics of Table AI. As unemployment insurance requires at least 0.5 years of paying premiums and gets paid longer as the premium-paying period gets longer, the average age of the C group is higher (35) than that of the T group (28). The T group is about one year more educated than the C group; for our later empirical analysis, six education dummies will be used. Table AI also shows the job categories of the last workplace. About 83% of the T group left their job voluntarily: 48% because of going self-employed, and 16% getting married or having a baby; reasons such as getting injured, being old or hitting the job age-limit are not entirely voluntary, although they are included in the 'voluntary quitting' category in the data source. The reason for voluntary quitting is not known for 16% of the T group ('personal reasons').

The row 'ex-firm size' shows the size (in persons) of the last workplace; 74% are under 500, but due to some outliers of size more than 10,000, the SD of ex-firm size is huge. The duration of work at the last workplace is 654 days for the T group on average, while it is 902 days for the C group. The duration between the time point of losing the job and enrolling in a job-training programme is about two months, whereas the duration between the time point of losing the job and starting to receive unemployment insurance is about one month. In terms of the industry to which the last workplace belonged (omitted in Table AI), manufacturing is leading with 38% (31%) for the T (C) group, followed by retail/wholesale with 16% (11%), real-estate rental with 15% (14%), and finance/insurance with 7% (11%); the other industry categories are health-related, construction, hotel/inn, transportation/warehouse, private education and public service.

Later, for estimation, we will drop the dummies for highschool (in education), machine operator (in job), job age-limit (in reason), manufacturing (in industry), and textile (in training type). Hence, the categories corresponding to the dropped dummies constitute the 'base case' unemployed woman.

Table I. Training types and durations

Training type	Number of trainees	Training duration
Textile	126	123
Machine/equipment	118	118
Information/communication	1151	107
Industrial application	274	150
Service industry	1728	138
Clerical/administration	820	108
Finance/insurance	88	74
Health-related	461	125

3. PROPORTIONAL HAZARD AND ACCELERATED FAILURE TIME

3.1. Proportional Hazard and Weibull MLE

Suppose $(t_1, q_1, x_1), \ldots, (t_N, q_N, x_N)$ are observed where t_i is a continuous unemployment duration, and q_i is a non-censoring indicator. As in Section 1, let $t_i = \min(y_i, c_i)$; ignore y_{ji} and d_i for a while. Let $t_1 \le t_2 \le \cdots \le t_N$ without loss of generality. 'Hazard' $\lambda(\tau, x)$ of leaving the unemployment state at time τ given x is defined as $\lambda(\tau, x) \equiv f(\tau|x)/S(\tau|x)$, where $f(y|x)$ is a density function for $y|x$ and $S(\tau|x) \equiv P(y > \tau|x)$ is the 'survival function'. Then

$$\lambda(\tau, x) = -\partial \ln S(\tau|x)/\partial\tau, \qquad S(\tau|x) = \exp\left[-\int_0^{\tau} \lambda(s, x)ds\right] = \exp(-\Lambda(\tau, x))$$

where $\Lambda(\tau, x) \equiv \int_0^{\tau} \lambda(s, x)ds$ is the 'integrated or cumulative hazard'. 'Proportional hazard' specifies $\lambda(\tau, x) = \lambda_o(\tau)\psi(x, \beta)$, where $\lambda_o(t)$ is the 'base-line hazard'. The word 'proportional' comes from the fact that the function $\psi(x, \beta)$ changes the base-line hazard proportionally.

If $y|x$ follows a Weibull distribution with parameters $(\beta_o, \theta(x))$, then $\lambda(\tau, x) = \beta_o \tau^{\beta_o - 1}\theta(x)$ and setting $\theta(x) = \exp(x'\beta)$ we get

$$\lambda(\tau, x) = \beta_o \tau^{\beta_o - 1}\exp(x'\beta) \quad (\Rightarrow \Lambda(\tau, x) = \tau^{\beta_o}\exp(x'\beta)) \tag{7}$$

which is a proportional hazard with $\lambda_o(\tau) = \beta_o \tau^{\beta_o - 1}$ and $\psi(x, \beta) = \exp(x'\beta)$. The role of a regressor, say z_i, with coefficient β_z is to multiply $\lambda_o(t)$ by $\exp(z_i\beta_z)$. The Weibull MLE maximizes

$$\sum_i \{q_i \ln f(t_i|x_i; b_o, b) + (1 - q_i)\ln S(t_i|x_i; b_o, b)\} = \sum_i \{q_i \ln \lambda(t_i|x_i; b_o, b) - \Lambda(t_i|x_i; b_o, b)\}$$

for b_o and b, which are for β_o and β. Later, for our data analysis, x_i will be replaced by x_i, d_i and interaction terms of elements in x_i and d_i; this will also be the case for the other estimators in this section.

3.2. Cox Partial Likelihood Estimator (PLE)

Weibull MLE is restrictive in a number of aspects. First, proportional hazard may not hold. Second, although specifying $\theta(x)$ as $\exp(x'\beta)$ is not controversial, specifying $\lambda_o(\tau)$ as $\beta_o\tau^{\beta_o - 1}$ is. Third, although irrelevant to our study, Weibull MLE is not flexible in allowing for time-varying regressors. The last two shortcomings are overcome in the Cox partial likelihood estimator (Cox, 1972).

Define (i) as the 'label' for the person failing ith, not being censored, $R(t_{(i)})$ as the 'risk set' at the ith failure (people surviving until just before the ith failure), and maintain the proportional hazard with the exponential function as in Weibull MLE: $\lambda(t, x_i(t)) = \lambda_o(t)\exp(x_i(t)'\beta)$, where x_i is allowed to depend on t. The risk set $R(t_{(i)})$ includes people who either failed or censored at or after $t_{(i)}$. The main idea of PLE rests on the likelihood of observing person i failing at t given a risk set $R(t)$ being

$$\lambda_o(t)\exp(x_i(t)'\beta)/\sum_{j\in R(t)}\{\lambda_o(t)\exp(x_j(t)'\beta)\} = \exp(x_i(t)'\beta)/\sum_{j\in R(t)}\exp(x_j(t)'\beta)$$

which is free of $\lambda_o(t)$.

Let M be the number of failure times (censoring times not counted). Since failure occurs only at $t_{(i)}$, the 'partial likelihood' function to maximize for b is

$$\prod_{i=1}^{M}\left\{\exp(x_{(i)}(t_{(i)})'b)\Bigg/\sum_{j\in R(t_{(i)})}\exp(x_j(t_{(i)})'b)\right\} \tag{8}$$

The partial likelihood may be taken as a 'regular' likelihood; the asymptotic variance can be estimated in the usual way for MLE.

3.3. Accelerated Failure Time (AFT)

In accelerated failure time, ln(duration) is specified as a linear function of x:

$$\ln y_i = x_i'\gamma + u_i$$

where γ is a parameter vector, u_i is independent of x_i and e^u has survival function S and cumulative hazard function Λ. Then, $y = e^{x'\gamma}e^u$ and

$$P(y > \tau|x) = P(e^u > \tau e^{-x'\gamma}|x) = S(\tau e^{-x'\gamma}) = \exp(-\Lambda(\tau e^{-x'\gamma}))$$

$e^{-x'\gamma}$ accelerates or decelerates τ, which explains the name AFT. AFT is a special case of 'transforming response', where a transformation of y_i other than $\ln(y_i)$ can appear on the left-hand side of the equation. But, as already mentioned, log-transformation renders a nice treatment effect interpretation, and we will not explore other transformations.

The AFT interpretation requires independence of u from x; otherwise, if u is related to x such that $e^u|x$ has survival function S_x and cumulative hazard Λ_x, then

$$P(y > \tau|x) = S_x(\tau e^{-x'\gamma}) = \exp(-\Lambda_x(\tau e^{-x'\gamma}))$$

x can influence the duration other than through $\tau e^{-x'\gamma}$.

If there is no censoring, we can estimate the AFT linear model simply with a least squares estimator (LSE) without specifying the error term distribution. To deal with the random censoring problem, however, we will assume $u \sim N(0, \sigma_u^2)$ independently of x for a constant $\sigma_u^2 > 0$. Then, we get an MLE maximizing, for g and s_u,

$$\sum_i\left(q_i\ln\left[\phi\left\{\frac{\ln(t_i) - x_i'g}{s_u}\right\}/s_u\right] + (1-q_i)\ln\left[1-\Phi\left\{\frac{\ln(t_i) - x_i'g}{s_u}\right\}\right]\right) \tag{9}$$

The AFT model has a number of advantages, compared with Weibull MLE and Cox PLE. First, since we have a familiar linear model, we can try many well-known specification tests. Second, suppose there is an unobserved variable, say v_i. Such a variable appears in the form $\exp(x_i'\beta + v_i)$ for Weibull MLE and PLE, which makes both inconsistent, even if v_i is independent of x_i. For AFT, v_i can be simply absorbed into u_i. With v in u, if there is no censoring, LSE for the AFT linear model is consistent so long as $\text{COR}(x, u) = 0$; if MLE is used for censoring, then independence of u from x (if not, a known form of dependence of u on x) is required. Thus, allowing for unobserved variables is much easier in AFT than in the hazard-based approaches; in the latter, various mixing distributions have been tried in the literature, but the resulting estimates are sensitive to the chosen distribution [see Van den Berg (2001) and references cited therein].

Let x_i include 1 as its first component as usual. Consider AFT models linear in x_i and w_i for y_{0i} and y_{1i}, where w_i consists of known functions of elements of x_i:

$$\ln(y_{0i}) = x_i'\beta_x + u_i, \qquad \ln(y_{1i}) = x_i'\beta_x + w_i'\beta_w + u_i, \qquad E(u_i|x_i, d_i) = 0$$
$$\Rightarrow \ln(y_i) = x_i'\beta_x + d_i w_i \beta_w + u_i, \qquad E(u_i|x_i, d_i) = 0$$

where β_x and β_w are unknown parameter vectors. Observe

$$E\{\ln(y_1) - \ln(y_0)|x\} = w'\beta_w \Rightarrow E\{\ln(y_1) - \ln(y_0)\} = E(w)'\beta_w$$

which are conditional and marginal treatment effects, respectively. If $w = 1$, then the conditional effect is the same as the marginal effect. Also, the conditional effect with $w = E(w)$ is equal to the marginal effect.

3.4. Empirical Evidence

Table AII shows the results of the Weibull MLE, Cox PLE and AFT. Examining the Weibull MLE and Cox PLE first, they are close in terms of signs, estimate magnitude and statistical significance. This means that the Weibull distributional assumption may not be so bad after all. Given the similarity and the fact that Cox PLE requires much weaker assumptions than Weibull MLE, we will interpret only Cox PLE in the following.

For exp(est.) in Table AII, consider a dummy variable for college graduation with coefficient 0.535. This means that the hazard out of unemployment for college graduates is exp(0.535) = 1.707 times the hazard for highschool graduates. For variables other than dummies, exp(est.) shows the same type of proportional effect as the variable increases by one unit from a 'reference level' (for dummies, the reference level is zero). Table AII clearly shows that there are many significant interaction terms. Unfortunately, interaction terms make the task of assessing covariate effects complicated. Interaction or not, stating that the hazard increases by a certain percentage is less palatable then stating that the duration increases by a certain percentage; we will be able to make the latter type of statements with AFT later.

For us, the most important variables are d and its interaction terms. The treatment d itself shows a negative effect: job training lengthens unemployment duration. It is possible, however, that this negative effect gets reversed when interaction terms are taken into account. Some possibilities are shown in Table II.

If an unemployed woman who worked in the transportation/warehouse industry is a professional (job category) and receives a finance/insurance training, then her hazard out of unemployment is 2.44 times greater than the base case (machine operator in the manufacturing industry receiving a textile training). If an unemployed woman who worked in the finance/insurance industry is a clerk (job category) and receives a finance/insurance training, then her hazard out of unemployment is 1.88 times greater than the base case.

In Table AII on AFT, all signs are opposite to those in Cox PLE, which is natural, because increasing hazard means decreasing duration. Age increases the duration: age and age × education carry positive numbers. Education decreases the duration as the dummies for junior college, college and graduate school indicate; the effect, however, is mitigated by age. Among the reasons for

Table II. Sub-populations with increased hazard out of unemployment

Sub-population	Effect estimate
prof., trans./ware., fin./ins.	$\exp(-1.334 + 0.459 + 0.808 + 0.960) = \exp(0.893) = 2.44$
clerk, fin./ins., fin./ins.	$\exp(-1.334 + 0.490 + 0.517 + 0.960) = \exp(0.633) = 1.88$

Table III. Sub-populations with decreased duration

Sub-population	Effect estimate
prof., trans./ware., fin./ins.	$1.829 - 0.433 - 0.827 - 1.013 = -0.444$
clerk, fin./ins., fin./ins.	$1.829 - 0.477 - 0.600 - 1.013 = -0.261$

Table IV. Sub-populations favourable for job training

Job	executives, (semi-)pro., clerical, service/sales
Industry	transportation/warehouse, fin./ins., private education, public service
Training type	machine/equipment, info./communication, clerical., fin./ins.

getting unemployed, 'self-employed' and 'marriage/baby' increase the duration. The size of the last workplace and the tenure there increase the duration. The unemployment duration before entering the data also increases the duration.

Turning to the treatment, d has a big positive number: for the base case, job training increases the duration, which corroborates the findings in Weibull MLE and Cox PLE. But when interactions are taken into account, we can find combinations of job, industry and training type with the opposite sign. Recalling the two sub-populations considered in Table II, their AFT effects are shown in Table III. This table shows a 44% and 26% reduction in the unemployment duration.

Overall, judging from the interaction terms, the job, industry and training type categories in Table IV seem to reduce the unemployment duration. It is interesting that while all advantageous job categories are white-collar jobs regardless of industry categories, advantageous training types are mixed. If one accepts the proposition that relatively more able workers are in white-collar jobs, then this finding means that more able workers get employed faster. The training types presumably reflect two things: one is the demand during the period 1999 to early 2000, and the other is the effectiveness of training.

4. MATCHING AND GROUP-MEAN DIFFERENCES

Call treated woman i simply 'case i'. In matching, the goal is to select one or more similar controls for case i. Sometimes this is done in two stages. In the first stage, some covariates are matched exactly, perhaps because they are deemed important; this is done by making strata with the covariates and considering only the controls in the same stratum as case i falls in. In the second stage, a distance measure is used to select controls closest to case i using the covariates not used in the first stage. If only the closest control is selected, we have a 'pair matching'; if more than one closest are selected, we have a 'multiple matching'. A popular distance measure is the 'Mahalanobis distance': for case i, it is

$$(x_i - x_m)' C_N^{-1} (x_i - x_m) \quad m \text{ indexing the controls in the same stratum}$$

where x is only for the second stage covariates and C_N is a sample covariance matrix for x using the C or T group.

It is possible that a case is passed over, if there is no reasonable match in the C group. If a control is used to match only one case (to simplify the ensuing statistical analysis), then call the matching 'greedy'; otherwise, if a control is allowed to be matched multiple times, the matching is 'non-greedy'. If we start matching with case 1 and then proceed to case 2 and so on, then the matching is 'sequential'; otherwise, if all cases are considered jointly, then the matching is 'non-sequential'. Clearly, non-sequential matching is better but hard to implement, for all $N_0 N_1$ pairs should be considered jointly; for non-greedy matching, unless there is a limit on the number of times a control can be matched, there is no reason to do non-sequential matching. Sequential greedy matching is particularly simple, for the 'control reservoir' keeps shrinking as the matching goes on.

Despite many different ways to do matching, getting the (asymptotic) variance for a treatment effect estimator with matching is not easy. For pair matching, the treatment effect estimator is $A_N \equiv N_u^{-1} \sum_{i \in T_u} (y_i - y_{mi})$, where y_{mi} is the matched control to case i, N_u is the number of used cases and T_u denotes the group of used cases (recall that some cases may not be used if no good match is found). Usually $y_i - y_{mi}$, $i \in T_u$, are taken as *iid*, and $V_N \equiv N_u^{-2} \sum_{i \in T_u} \{(y_i - y_{mi}) - A_N\}^2$ is used as an estimator for the asymptotic variance. In principle, however, this variance estimator needs a justification, because all controls are involved in matching, which implies dependence across the pairs in the sum.

Pair matching is an extremely simple case. At the other end of the spectrum, there is (purely) nonparametric matching using a weighted average of all controls for each case; the asymptotic distribution is available in Heckman *et al*. (1998) when nonparametric kernel methods are used. For the intermediate cases, there is no general asymptotic variance formula; each matching estimator calls for the derivation of its own asymptotic variance, which is by no means easy.

Given the huge sample size in our data, we will use two pair-matching schemes that are easy to implement. One is sequential greedy, and the other is sequential non-greedy; the only difference is that the latter allows a control to be matched multiple times. The former is simpler, but the latter tends to balance covariates better at the cost of a higher SD, as Abadie and Imbens (2002) state. In sequential non-greedy pair matching, however, multiple use of the same control units would make it harder to justify the asymptotic variance estimator V_N. Abadie and Imbens (2002) derive the asymptotic distribution for sequential non-greedy matching when x is continuous and the number of matched controls is identical for all cases.

In our x, most variables are dummies; these dummies were matched exactly. Among the non-dummy variables, six dummies were used for education and ex-firm size. For education, the six categories are primary school graduation, middle school, high school, junior college, college and graduate school; for ex-firm size, 1–10, 11–50, 51–100, 101–300, 300–500 and 500$^+$ (the last category 500$^+$ includes large conglomerates of size 10,000 or above). These dummies for education and ex-firm size were also matched exactly.

For given strata determined by the dummies, age and job experience at the last workplace were used for Mahalanobis distance. We also included the duration between the time point of losing the job and enrolling in job training (or receiving unemployment insurance) in Mahalanobis distance,

Table V. Matching evaluation

	After matching						Before matching		
	Greedy			Non-greedy			T	C	t-Value
	T	C	t-Value	T	C	t-Value			
Age	27.7	28.4	−6.1	27.8	28.1	−2.9	27.8	34.9	−76.8
Job exp. (yrs)	1.83	1.89	−1.4	1.79	1.76	0.9	1.79	2.47	−21.4
Last unemp. dur. (days)	65.1	53.3	11.0	65.0	58.1	7.1	65.1	33.6	41.1

because this duration may alter the reservation wage of the unemployed: a longer duration may frustrate the woman and make her more willing to take a given job offer. Also, for inequality (6), the time point at which one enrolls in a job-training programme or starts to receive unemployment insurance was matched: we counted dates with 1 January 1999 as 1 until the end of March 2000, and the date variable was used also for Mahalanobis distance.

The two pair-matching schemes are evaluated in Table V by comparing the mean values of the covariates used for Mahalanobis distance. Non-greedy matching does slightly better than greedy matching. Although there are still some differences left after matching between the two groups, the differences are much smaller compared with before matching and seem to be negligible economically.

All but the last rows of Table AIII show the treatment effects (6) in the left half and (5) in the right half for various sub-populations listed in the first column. The right half shows that the treatment increases the duration for all sub-populations considered; from the last row, the marginal (or average) effect is about 40 days, which is significant. But as already mentioned, this finding does not use the censored observations.

Turning to the left half of Table AIII, there, all censored observations are used. For most sub-populations, the employment effect is negative, ranging approximately from insignificant −0.04 (health training) to significant −0.15 (service training). But there are also some positive effects, ranging approximately from significant 0.04 (information/communication training) to significant 0.15 (finance/insurance training). Recall that the training duration for finance/insurance training is only 74 days, which is much shorter than the average training duration of about 120 days; this may explain why finance/insurance training was successful. For the entire population, the last row for the marginal effect shows significant negative numbers −0.064 and −0.073: job training decreases employment probability by about 6–7%.

5. CONCLUSIONS

In this paper, we assessed job-training effects on Korean women for the period January 1999 to March 2000. We used a number of different measures of job-training success, and different estimation methods accordingly. Despite some discrepancies across the results using the methods, clear common findings emerged: overall, job training increased unemployment duration and, as such, needs fixing. Looking at the reason why, job training seemed too long relative to its effectiveness in placing the trainees. It is, however, possible that job training may contribute by extending the future employment duration, once employed; but to check this out, more data are needed. Despite the overall gloomy picture painted in this paper on job-training effects, there are some sub-groups for which job training decreased unemployment duration: for ex-job types, they are white-collar workers regardless of industries; for training types, finance/insurance and information/communication. At least, this shows which groups of ex-job and training types the government should target to improve job-training outcomes.

APPENDIX

Table AI. Summary statistics

	Treatment group		Control group		*t*-Value
	Mean	SD	Mean	SD	
Age (years)	27.789	5.566	34.924	10.815	−76.75
Education (years)	13.070	1.760	12.128	2.505	34.42
Job: executive	0.003	0.051	0.010	0.101	−9.13
Job: professional	0.034	0.182	0.040	0.195	−2.05
Job: semi-professional	0.063	0.243	0.057	0.232	1.56
Job: clerical	0.458	0.498	0.505	0.500	−6.33
Job: service/sales	0.144	0.351	0.102	0.303	8.17
Job: mechanic	0.156	0.363	0.104	0.306	9.62
Job: machine operator	0.018	0.132	0.014	0.117	1.97
Job: menial labour	0.124	0.330	0.167	0.373	−8.53
Voluntary quitting	0.825	0.379	0.084	0.278	134.84
Reason: self-employed	0.477	0.500	0.022	0.146	64.34
Reason: marriage/baby	0.160	0.367	0.014	0.117	28.11
Reason: injured/old	0.035	0.184	0.030	0.171	1.51
Reason: personal	0.156	0.362	0.004	0.064	29.52
Reason: job age-limit	0.000	0.000	0.015	0.120	−26.42
Ex-firm size (persons)	1070	3760	840	3277	2.23
Job experience at ex-firm (days)	653.9	806.6	901.5	481.1	−21.4
Unemployment duration before enrollment (days)	65.1	53.2	33.6	32.5	41.1

Table AII. Weibull MLE, Cox PLE and AFT

	Weibull MLE		Cox PLE			AFT	
	est.	tv	est.	tv	exp(est.)	est.	tv
1	−4.896	−60.75	–	–	–	4.159	47.59
Age	−1.590	−2.24	−1.566	−2.21	0.548	2.785	3.99
Primary school	−0.764	−3.97	−0.748	−3.89	0.473	0.457	2.44
Middle school	−0.168	−1.88	−0.160	−1.80	0.852	0.036	0.41
Junior college	0.461	11.53	0.452	11.33	1.571	−0.396	−9.74
College	0.547	7.45	0.535	7.30	1.708	−0.419	−5.71
Graduate school	0.797	5.01	0.792	4.98	2.207	−0.645	−3.81
Age × edu	−0.235	−3.95	−0.229	−3.86	0.725	0.128	2.20
Job: executive	−0.589	−5.29	−0.585	−5.26	0.557	0.696	6.00
Job: professional	−0.571	−7.38	−0.568	−7.35	0.566	0.602	7.06
Job: semi-professional	−0.476	−6.46	−0.481	−6.54	0.618	0.553	6.79
Job: clerical	−0.567	−8.66	−0.569	−8.70	0.566	0.610	8.34
Job: service/sales	−0.684	−9.49	−0.688	−9.57	0.503	0.741	9.45

Table AII. (*Continued*)

	Weibull MLE		Cox PLE			AFT	
	est.	tv	est.	tv	exp(est.)	est.	tv
Job: mechanic	−0.432	−6.35	−0.428	−6.29	0.652	0.422	5.63
Job: menial labour	−0.129	−1.94	−0.137	−2.05	0.872	0.156	2.12
Reason: self-employed	−0.191	−3.22	−0.157	−2.64	0.855	0.209	3.42
Reason: marriage/baby	−0.514	−5.62	−0.506	−5.54	0.603	0.462	5.49
Reason: injured/old	−0.021	−0.38	−0.031	−0.56	0.970	0.053	0.97
Reason: personal	0.231	1.91	0.230	1.91	1.259	−0.252	−1.91
Ex-firm size	−0.063	−11.46	−0.060	−11.08	0.984	0.049	12.08
Job experience	−0.044	−6.19	−0.041	−5.73	0.897	0.043	5.81
Pre-unemployed duration	−0.004	−11.75	−0.004	−13.11	0.876	0.004	13.15
Industry: construction	−0.016	−0.46	−0.020	−0.58	0.980	0.039	1.02
Industry: retail/wholesale	−0.142	−4.64	−0.151	−4.93	0.860	0.156	4.84
Industry: hotel/inn	−0.398	−6.45	−0.409	−6.63	0.665	0.358	6.20
Industry: trans./warehouse	−0.841	−17.74	−0.825	−17.55	0.438	0.861	20.78
Industry: finance/ins.	−0.645	−16.52	−0.638	−16.44	0.528	0.674	17.94
Industry: real-estate rental	0.111	4.37	0.104	4.10	1.110	−0.116	−4.18
Industry: private education	−0.679	−11.84	−0.679	−11.90	0.507	0.697	12.43
Industry: health-related	−0.159	−4.00	−0.162	−4.08	0.850	0.156	3.71
Industry: public service	−0.383	−7.05	−0.383	−7.06	0.682	0.397	7.27
d	−1.365	−3.15	−1.334	−3.08	0.263	1.829	4.17
$d \times$ (job: executive)	0.999	1.76	0.976	1.72	2.654	−0.909	−1.48
$d \times$ (job: professional)	0.455	1.54	0.459	1.56	1.583	−0.433	−1.44
$d \times$ (job: semi-professional)	0.383	1.40	0.393	1.44	1.482	−0.403	−1.45
$d \times$ (job: clerical)	0.482	1.92	0.490	1.95	1.632	−0.477	−1.90
$d \times$ (job: service/sales)	0.445	1.67	0.452	1.70	1.571	−0.437	−1.65
$d \times$ (job: mechanic)	0.262	1.01	0.262	1.01	1.299	−0.236	−0.91
$d \times$ (job: menial labour)	0.098	0.38	0.113	0.44	1.120	−0.119	−0.46
$d \times$ (industry: construction)	0.067	0.50	0.072	0.54	1.074	−0.059	−0.40
$d \times$ (industry: retail/wholesale)	0.045	0.46	0.052	0.53	1.053	−0.065	−0.63
$d \times$ (industry: hotel/inn)	0.232	1.15	0.241	1.19	1.273	−0.230	−1.13
$d \times$ (industry: trans./warehouse)	0.826	4.77	0.808	4.66	2.243	−0.827	−4.46
$d \times$ (industry: finance/ins.)	0.526	3.86	0.517	3.79	1.676	−0.600	−4.41
$d \times$ (industry: real-estate rental)	−0.066	−0.77	−0.064	−0.75	0.938	0.051	0.54
$d \times$ (industry: private education)	0.368	2.33	0.380	2.41	1.463	−0.475	−2.93
$d \times$ (industry: health-related)	0.076	0.54	0.083	0.59	1.086	−0.123	−0.86
$d \times$ (industry: public service)	0.333	1.89	0.336	1.91	1.400	−0.456	−2.50
$d \times$ (training: machine/equipment)	0.598	2.34	0.579	2.27	1.785	−0.675	−2.61
$d \times$ (training: info./commun.)	0.609	2.83	0.590	2.74	1.804	−0.608	−2.95
$d \times$ (training: industrial application)	0.001	0.00	−0.001	0.00	0.999	0.028	0.12
$d \times$ (training: service)	−0.294	−1.35	−0.292	−1.34	0.747	0.249	1.21
$d \times$ (training: clerical)	0.485	2.22	0.465	2.13	1.592	−0.497	−2.38
$d \times$ (training: finance/ins.)	0.998	3.77	0.960	3.63	2.612	−1.013	−3.68
$d \times$ (training: health)	0.281	1.24	0.277	1.22	1.320	−0.281	−1.28
$d \times$ (reason: self-employed)	0.263	2.74	0.221	2.31	1.248	−0.244	−2.40
$d \times$ (reason: marriage/baby)	0.012	0.08	0.003	0.02	1.003	0.015	0.11
$d \times$ (reason: injured/old)	−0.080	−0.44	−0.074	−0.40	0.929	0.018	0.10
$d \times$ (reason: personal)	−0.262	−1.71	−0.267	−1.74	0.766	0.292	1.76
$d \times$ (ex-firm size)	0.067	8.05	0.064	7.78	1.003	−0.052	−6.21
$d \times$ (job experience)	−0.046	−2.30	−0.047	−2.38	0.886	0.030	1.65
$d \times$ age	−1.199	−1.74	−1.200	−1.74	0.631	−0.013	−0.02
$d \times$ edu	0.040	2.19	0.038	2.11	1.012	−0.037	−1.90
$d \times$ (pre-unemployed duration)	0.003	4.27	0.003	4.43	1.077	−0.003	−4.69
α	0.931	–	–	–	–	–	–

Table AIII. Treatment effects with matching

	Employment-or-not-effect				Duration effect (days)			
	Greedy		Non-greedy		Greedy		Non-greedy	
	est.	tv	est.	tv	est.	tv	est.	tv
Middle school	−0.175	−3.622	−0.139	−3.036	0.000	0.000	0.000	0.000
High school	−0.061	−5.398	−0.083	−7.385	43.5	9.148	52.6	11.250
Junior college	−0.074	−3.527	−0.069	−3.374	51.4	8.434	47.8	8.079
College	−0.002	−0.102	−0.005	−0.252	29.5	4.548	32.7	5.215
Job: professional	−0.047	−0.988	−0.053	−1.120	31.2	1.654	38.0	2.084
Job: semi-professional	−0.051	−1.314	−0.074	−2.007	20.3	1.688	29.0	2.583
Job: clerical	−0.061	−4.490	−0.048	−3.598	45.8	10.60	47.5	11.14
Job: service/sales	−0.053	−2.443	−0.057	−2.676	39.2	4.104	438	4.724
Job: mechanic	−0.031	−1.343	−0.110	−4.995	47.3	4.665	58.5	6.169
Job: machine operator	−0.145	−2.257	−0.048	−0.759	43.1	1.248	38.3	1.107
Job: menial labour	−0.069	−2.775	−0.107	−4.306	39.3	4.038	46.1	4.781
Training: textile	−0.079	−1.558	−0.143	−2.671	48.8	1.865	48.8	1.865
Training: machine/equip.	0.051	0.869	0.051	0.869	34.7	1.825	34.7	1.825
Training: info./commun.	0.049	2.525	0.037	1.908	34.8	6.372	34.7	6.398
Training: indus. appl.	−0.107	−2.818	−0.099	−2.643	80.4	7.092	83.9	7.358
Training: service	−0.141	−10.16	−0.158	−11.36	64.9	8.979	71.1	9.881
Training: clerical	−0.044	−1.928	−0.046	−2.097	22.7	3.417	29.4	4.476
Training: finance/ins.	0.148	2.028	0.159	2.166	6.6	0.459	6.641	0.459
Training: health	−0.038	−1.295	−0.047	−1.583	46.8	4.165	48.1	4.264
Marginal effect	−0.064	−20.66	−0.073	−23.57	38.8	2.894	44.2	9.164

ACKNOWLEDGEMENTS

We are grateful to a referee for his/her detailed comments.

REFERENCES

Abadie A, Imbens GW. 2002. Simple and bias-corrected matching estimators for average treatment effects. NBER Technical Working Paper 283.

Angrist J, Krueger A. 1999. Empirical strategies in labor economics. In *The Handbook of Labor Economics III*, Ashenfelter O, Card D (eds). North-Holland: Amsterdam.

Cox DR. 1972. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**: 187–220.

Heckman JJ, Ichimura H, Todd P. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* **65**: 261–294.

Heckman JJ, LaLonde R, Smith J. 1999. The economics and econometrics of active labor market programs. In *The Handbook of Labor Economics III*, Ashenfelter O, Card D (eds). North-Holland: Amsterdam.

Lee MJ. 2000. Median treatment effect in randomized trials. *Journal of the Royal Statistical Society, Series B* **62**: 595–604.

Lee MJ, Kobayashi S. 2001. Proportional treatment effects for count response panel data: effects of binary exercise on health care demand. *Health Economics* **10**: 411–428.

Meyer BD. 1990. Unemployment insurance and unemployment spells. *Econometrica* **58**: 757–782.

Røed K, Zhang T. 2003. Does unemployment compensation affect unemployment duration. *Economic Journal* **113**: 190–206.

Van den Berg GJ. 2001. Duration models: specification, identification and multiple durations. In *The Handbook of Econometrics V*, Heckman JJ, Leamer E (eds). North-Holland: Amsterdam.

